Integration and Implementation Insights

Research resources for understanding and acting on complex real-world problems

What every interdisciplinarian should know about p values

☐ April 30, 2019May 14, 2019   ☐ Community Member   ☐ Concepts   ☐ Author - Alice Richardson, Complexity, Interdisciplinarity, Partner - PopulationHealthXchange, Uncertainty

Community member post by Alice Richardson

In interdisciplinary research it's common for at least some data to be analysed using statistical techniques. Have you been taught to look for 'p < 0.05' meaning that there is a less than 5% probability that the finding occurred by chance? Do you look askance at your statistician colleagues when they tell you it's not so simple? Here's why you need to believe them.

The whole focus on $p < 0.05$ to the exclusion of all else is a historical hiccup, based on a throwaway line in a manual for research workers. That manual was produced by none other than R.A. Fisher, giant of statistical inference and inventor of statistical methods ranging from the randomised block design to the analysis of variance. But all he said was that "[p = 0.05] is convenient to take … as a limit in judging whether a deviation is to be considered significant or not." Convenient, nothing more!

Looking solely for $p < 0.05$ and deciding that the result is significant or not has the effect of replacing a number that can range between 0 and 1 with a binary decision (significant or not significant). This is a waste of information, an inefficient use of experimental resources.

There's definitely a feeling amongst statisticians that researchers need to embrace a world beyond $p < 0.05$. In 2016 the American Statistical Association published a statement on p values (Wasserstein and Lazar 2016). Its aim was to alert the research community to the problems associated with an over-reliance on $p < 0.05$, and to propose some principles for future research to follow.

However, this statement did not mark the end of a debate. The American Statistical Association has published again (Wasserstein, Schirm and Lazar 2019), and not just the editorial cited but a collection of over 40 individual articles. This collection encompasses everything from the history of the p value debate through further alternatives to the p value, such as effect sizes, Bayesian factors and so on to changing the balance in statistics education.

Alice Richardson (biography (https://i2insights.org/20 p-values/#alice-richardson))

The science press has picked up on the collection, with a Nature article (Arnheim, Greenland and McShane 2019) attracting the signatures of over 800 eminent statisticians and scientists who are keen to see continued reduction in the weight attached to $p < 0.05$.

Statisticians are thinking hard about how to do this. Educators are calling for revisions to standard introductory statistics courses to emphasise statistical thinking. Some are taking a hard line, such as the Journal of Basic and Applied Psychology in 2016 banning the publication of p-values in their journal, as was widely reported in the scientific press at the time. It makes me feel as though every research department should have a sign over the door saying:

> "Abandon Statistical Significance, all Ye who Enter Here!"

My thoughts on the way forward revolve around two concepts central to considering complexity:

- embracing uncertainty; and,
- thinking critically.

These ideas are hardly new – interdisciplinarians and other researchers have been advocating for this for years, and my view is that now is the time for these practices to become second nature.

Asking questions like: So what? Compared to what? How precise are the estimates? What was the model? Are assumptions of independence and random sampling likely to have been met? How robust are the results to changes in or departures from the model? Being transparent about responses to these are the way that science will advance.

Researchers have been painted into a corner where the maxims of "publish or perish" and "p < 0.05 or it's not publishable" drive the research agenda. It's not going to be easy to move to a new world, which is where I think complexity scientists come in.

How do you think we could progress these changes? Are there ideas that statisticians could learn from complexity scientists? How would you encourage moves towards embracing uncertainty and thinking critically?

References:
Arnheim, V., Greenland, S. and McShane, B. (2019). Scientists rise up against statistical signficance. *Nature*, 576: 305–307.

Wasserstein R.L., and Lazar, N.A. (2016). The ASA's statement on p-values: context process and purpose. *American Statistician*, **70**: 129-133.

Wasserstein, R.L., Schirm, A.L. and Lazar, N.A. (2019). Moving to a world beyond 'p < 0.05'. *American Statistician*, **73**, supp1: 1–19.

Biography: *Alice Richardson PhD is a biostatistician in the National Centre for Epidemiology and Population Health, Research School of Population Health at The Australian National University (ANU) in Canberra, Australia. Prior to commencing at ANU she taught introductory statistics at the University of Canberra for twenty years, providing a wealth of opportunities for communicating the complexities of "p < 0.05" to a diverse audience. Her research now focuses on imputation of missing data in highly structured data sets in order to extract maximum value from complex data collections.*

Alice Richardson is a member of blog partner PopulationHealthXchange (https://i2insights.org/about/partners#pophealthxchange), which is in the Research School of Population Health at The Australian National University.

7 thoughts on "**What every interdisciplinarian should know about p values**"

1. **josephguillaume** says:
   April 30, 2019 at 10:17 am

   Thanks for the nice summary and call for embracing uncertainty.

   I would argue that it's horses for courses, and that 0.05 still has value precisely in the conditions it was originally proposed: when a convenient cut-off is needed to make a binary decision about whether an event may have occurred by chance.

   I do have a problem with the statement that "Looking solely for $p < 0.05$ and deciding that the result is significant or not has the effect of replacing a number that can range between 0 and 1 with a binary decision (significant or not significant). This is a waste of information, an inefficient use of experimental resources."

   Not including the p value itself does throw away information, but this is often precisely what is needed for effective communication. Rather than forcing the decision maker (e.g. reader) to deal with probabilities themselves, the author can do some of the work for them. For example, if a statistical analysis has determined a p-value that a mechanical component is about to fail, the author might use a risk analysis to determine whether that p-value warrants action. This is also a binary cut-off, just based on more than only convenience.

   In my own work, I've tried to explore the variety of ways uncertainty is approached in different contexts, especially in how it is communicated (or "framed").
   Guillaume JHA, Helgeson C, Elsawah S, Jakeman AJ, Kummu M (2017) Toward Best Practice Framing of Uncertainty in Scientific Publications: A Review of Water Resources Research Abstracts. Water Resources Research, July. https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2017WR020609 (https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2017WR020609)
   I'd be very interested in feedback from anybody interested in this topic.

   As you say, let's embrace uncertainty – in all its forms

   Reply (/2019/04/30/replacing-p-values/?replytocom=23199#respond)
   - **aliceinstatisticsland** says:
     May 1, 2019 at 3:45 pm

     Thanks for your thoughts Joseph. I do recognise that while it is wasteful to reduce a p value to a binary decision but that there can be circumstances where that binary decision is exactly what is required! Embracing uncertainty should be allowed to extend to the "horses for courses" you mention, providing the right information for an informed decision. In every setting.

     I enjoyed reading your article on uncertainty framing. Its conclusions and advice will certainly make me think carefully before dashing off an abstract to toss in at the front of a journal article!

     Reply (/2019/04/30/replacing-p-values/?replytocom=23288#respond)
2. **Jacques Boulet** says:
   April 30, 2019 at 10:16 am

   Phew!!!! finally uncertainty and critical thinking are 'allowed' (again)… and serious consideration of probability, plausibility, possibility in a complex and hyper-interactive and reciprocally interconnecting world becoming the primary consideration… so additional question(s) would pertain to the effects of reducing such complexities to 'single' variables and factors extracted from their 'ongoingness' in the complexity and attempting to replace them by a set of 'arrows' of assumed effect directionality (or 'causality'); the relational effects of the 'observer' – 'observed' and our assumptions about this; the ways in which we 'dialectically' deal with the time/place/ongoingness conundrum… (Bohm); etc. Thanks for bringing this all up again… brings joy to my old days… Jacques Boulet

   Reply (/2019/04/30/replacing-p-values/?replytocom=23198#respond)
   - **aliceinstatisticsland** says:
     May 1, 2019 at 3:42 pm

     Thanks Jacques! The issue you raise around measurement is similar to Kirsten, with the extra layer of consideration of how to reduce a large number of variables to 'single' variables. Causality is a topic of immense interest at the moment as well, somewhat on the edge of my expertise and so I'd be interested to hear others' views on how it fits in here.

     Reply (/2019/04/30/replacing-p-values/?replytocom=23287#respond)
3. **Kirsten Kainz** says:
   April 30, 2019 at 7:14 am

   I really appreciate this blog post! Thanks for writing. To add to your list of issues for consideration: how trustworthy are the measurement scheme and resulting data? What is the statistical model and how well does the model correspond with deep understanding of the problem? What can be learned by positing and testing multiple statistical models of the data (representing different perspectives on the problem and different theoretical orientations)? It would be interesting to work across nations and disciplines using common data to explore the types of questions and answers produced through diverse perspectives and models.

   Reply (/2019/04/30/replacing-p-values/?replytocom=23191#respond)

- **aliceinstatisticslandALice Richardson** says:
  May 1, 2019 at 9:02 am

  Thanks Kirsten! I liked your extra question about measurement as it is all very well to build a really complicated statistical model but the values that are fed into it have to be believable. My favourite quote on this topic comes from Josiah Stamp: "The government are very keen on amassing statistics. They collect them, add them, raise them to the nth power, take the cube root and prepare wonderful diagrams. But you must never forget that every one of these figures comes in the first instance from the village watchman, who just puts down what he damn pleases."

  Testing multiple statistical models is also a really good idea that usually gets pushed to one side because there just aren't the resources to do so (both in terms of time and skills). There is a recent example where the resources were brought together: 29 teams involving 61 analysts used the same data set to address a research question about racial bias in the awarding of red cards in soccer (football). Despite the apparent simplicity of the question, there were multiple approaches and multiple answers!

  References

  Silberzahn, R., Uhlmann, E. L., Martin, D. P., Anselmi, P., Aust, F., Awtrey, E., . . . Nosek, B. A.
  (2018). Many analysts, one data set: Making transparent how variations in analytic choices affect results. Advances in Methods and Practices in Psychological Science, 1, 337–356. doi:10.1177/2515245917747646.
  Stamp, Josiah (1929). Some Economic Factors in Modern Life. P. S. King & Son. pp. 258–259.

  Reply (/2019/04/30/replacing-p-values/?replytocom=23272#respond)

  - **Kirsten** says:
    May 3, 2019 at 10:53 pm

    Thanks for the Sizberzahn et al. reference. Printing it now!!!

    Reply (/2019/04/30/replacing-p-values/?replytocom=23448#respond)

This site uses Akismet to reduce spam. Learn how your comment data is processed (https://akismet.com/privacy/).

- **aliceinstatisticslandALice Richardson** says:
  May 1, 2019 at 9:02 am

  Thanks Kirsten! I liked your extra question about measurement as it is all very well to build a really complicated statistical model but the values that are fed into it have to be believable. My favourite quote on this topic comes from Josiah Stamp: "The government are very keen on amassing statistics. They collect them, add them, raise them to the nth power, take the cube root and prepare wonderful diagrams. But you must never forget that every one of these figures comes in the first instance from the village watchman, who just puts down what he damn pleases."